

Protein Spot Detection by Symmetry Derivatives of Gaussians

Martin Persson and Josef Bigun

School of Information Science, Computer
and Electrical Engineering, Halmstad University
Box 823, SE-301 18 Halmstad, Sweden
Martin.Persson@ide.hh.se

Abstract. Two-dimensional gel electrophoresis is the preferred method for simultaneously separating and visualising thousands of proteins. An important part of the computer aided analysis of the proteome is the ability to automatically detect, identify, and quantify the proteins by means of automatic Image processing. We present a fast and sensitive method for protein spot detection using the Circular Symmetry Tensor. It is based upon the work of Bigun on Symmetry derivatives of Gaussians.

1 Introduction

Two-Dimensional gel electrophoresis (2-DE) [1] is a powerful method for separating thousands of proteins within a biological sample. The samples are broken down into a solution of individual proteins, which is loaded onto an immobilised pH gradient (IPG) gel strip. The strip is subjected to an electric field, separating the proteins by their isoelectric point (isoelectric focusing, IEF). The separated strip is then placed on top of a SDS-PAGE gel and subjected to a second electric field, perpendicular to the original strip, resulting in separation by molecular weight in the second dimension. The resulting protein map can be visualised as a spot pattern (figure 1) after the gel has been subjected to a staining procedure.



Fig. 1. A 2-DE image of *Escherichia Coli*

Protein spot patterns produced by 2-DE are used to find changes in protein expression induced in the sample under different circumstances, identification of proteins by matching against model organism with known protein expressions, as well as isolating target proteins for further analysis by Mass Spectroscopy (MS). These applications require automatic identification and accurate quantification of relevant proteins by image analysis. The task is however hampered by the poor reproducibility of the 2-DE process. Artifact streaks, non-linear geometric deformations due to unevenly dried gels, saturated spots, and large background variations are some examples of distortions that frequently occur in 2-DE gels.

Visual inspection of 2-DE gels utilise the fact that proteins generally appear as peaks in roughly circular spots. We propose the use of a point of interest detector based upon Symmetry derivatives of Gaussians[2] for protein spot detection.

2 Symmetry derivatives of Gaussians and Circular Symmetry

Symmetry Derivatives are differential operators defined as [2]:

$$D_x + iD_y = d/dx + id/dy \quad (1)$$

which yield a complex vector field when applied to 2-D signals. Higher order symmetry derivatives and their conjugates are defined as

$$(D_x + iD_y)^n \quad (2)$$

$$(D_x - iD_y)^n \quad (3)$$

respectively. These are referred to as n'th symmetry derivative and n'th conjugate symmetry derivative respectively. When applied to an orientation tensor field [3] image $z(x,y)$ the filter responses can be interpreted as *linear symmetry* [3] after a coordinate transformation [4]. In particular the 2nd conjugate symmetry derivative cor-

responds to linear symmetry in a polar coordinate system, which gives a measure of *Circular Symmetry (CS)*.

The filter used for detection of Circular Symmetry is the complex exponential $\exp(-2i\phi)$. We choose to use the discrete approximation of $h=(x - iy)^2g(x,y)$ of this filter, where g is a 2D Gaussian. A Gaussian is chosen because it has two properties that are attractive in our implementation, namely separability and isotropy.

The phase of the complex filter response represents the "direction" in the polar coordinate system (figure 2) while the magnitude is a certainty measure of the direction. The local orientation (in the polar coordinate system) at the center of a protein spot will be dominantly radial.

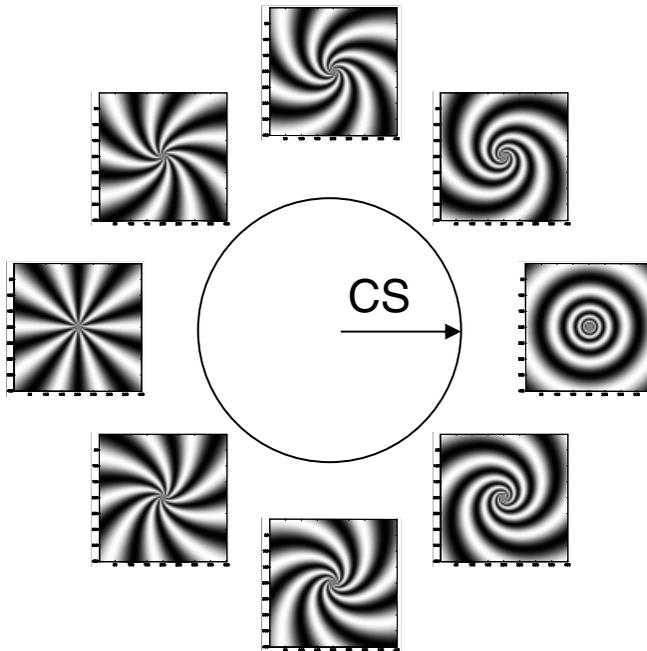


Fig. 2. The relationship between the argument of the Circular Symmetry tensor and symmetric functions in the polar coordinate system.

3 Circular Symmetry applied to protein spot detection

3.1 Implementation

The Circular Symmetry tensor for a neighbourhood in the image can be formulated as the scalar product $\langle h, z \rangle$. Here h is the complex filter defined as $h = (x - iy)^2g(x,y)$ while z is the local orientation (tensor) image computed as $z(x, y) = (dfx + idfy)^2$.

The scalar product is implemented as a 2D convolution between z and h since it is necessary to compute the scalar product for every neighbourhood. Since we are using 2D Gaussians we can take advantage of their separable property for a faster implementation. Essentially only two 1D filters are required for this step, $g(t)$ and $tg(t)$. Transposed filters are used in the other dimension. The pseudo code for the computation is given in table 1.

Table 1. Circular Symmetry from z . Here $*$ represents convolution and i represents $(-1)^{1/2}$

$$\begin{aligned}
 \text{tmp} &\leftarrow g^{\top} * (tg * R[z]) + g * (tg^{\top} * I[z]) + i (g^{\top} * (tg * I[z]) - \\
 &\quad g * (tg^{\top} * R[z])) \\
 \\
 \text{CS} &\leftarrow g^{\top} * (tg * R[\text{tmp}]) + g * (tg^{\top} * I[\text{tmp}]) \\
 &\quad + i (g^{\top} * (tg * I[\text{tmp}]) - g * (tg^{\top} * R[\text{tmp}]))
 \end{aligned}$$

The resulting image CS is a vector image where the argument encodes not only circular patterns but also radial patterns (figure 2). Only arguments close to 0 are of interest in this application. Therefore a threshold was applied to $\cos(\arg(CS))$.

Since the orientation tensor uses double the angle of the local gradient an inverse spot (figure 3), ie a region that is whiter than the surrounding area will result in the same orientation tensor as a dark spot on brighter background. We avoid this problem by thresholding the Laplacian (Lpf) of the 2-DE image.

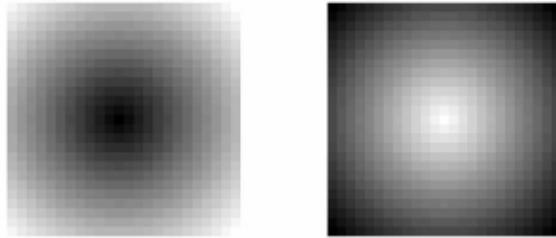


Fig. 3. True and false spots

Both the Laplacian Lpf and the local orientation tensor image z are computed using 1D Gaussians and their derivatives in a similar way to the computation of CS .

3.2 Experimental setup

We used two quantified standard gels from a comparison of 2-DE image analysis software previously conducted by Raman, Cheung, and Marten[5][6] in our evaluation of the algorithm. The gels in question are reference gels provided by the manufactur-

ers of two commercial 2-DE image analysis software's, Z3 and Melanie respectively. The gel images are in uncompressed 8-bit greyscale TIFF format and are of sizes 666 x 727 and 998 x 1087 pixels respectively. Evaluation of results was done by comparison with the reference images from [5] and [6]. Protein spots are manually counted in the reference images.

The orientation tensor image z was computed with 1D Gaussians of $\sigma=0.5$.

In order to increase the effect of small gradients (e.g. spots in streaks, cluttered spots, weak spots) we also use a rescaled z , defined as, $zmagn = z^{0.10}$.

Circular Symmetry detection is done by using 1D Gaussian filters of $\sigma=1.3$. The chosen values of σ are small due to the fine nature of the protein spots in our test images. From the Circular Symmetry image CS we compute an image $absCS$ containing the absolute values from elements in CS with sufficiently circular symmetry according to

$$\cos(CS) > 0.95 \Rightarrow absCS = |CS| \quad (4)$$

$$\cos(CS) < 0.95 \Rightarrow absCS = 0$$

In order to identify false "white" spots we compute the Laplacian for the original image. Candidate regions with $Lpf > -eps$ and $absCS > 0.5$ are accepted as spots.

3.3 Experimental results

Table 2. Results. Raman et al manually identified 955 spots in Gel a and 1403 spots in Gel b

	Gel a No	Gel b No
True Positives	865	1375
False Positives	321	609
False Negatives	90	28
FP/TP	37.1	44.29
FN/TP	10.4	2.04

The high number of False Positives stems mainly from two sources. Over segmentation (induced by the Laplacian threshold as well as the σ values being too small) and isolated noise artefacts of approximately circular shape. The False negatives are either the result of cluttered spots or of large, saturated spots with no gradients at the chosen scale. As a comparison Raman et al [5] achieved FN/TP of less than 11% with Z3 and between 7 and 13% with Melanie 3.0. Their FP/TP was ~10% for Z3 and 35-350% for Melanie (depending on sensitivity parameters).

4 Conclusion

Given a crude classification scheme Circular Symmetry yields results comparable to commercial software packages. We find this fact very promising in the search for an automatic POI extractor for 2-DE images.

The misclassifications are mainly caused by the choice of scale. The fine scale chosen results in increased sensitivity to minute specks and other noise on the gel, as well as over segmentation of large spots. These problems are well known within both the pattern recognition and the 2-DE community. We expect to improve the algorithm by adding multiscale spot detection and region merging to help with the over segmentation problem.

It is also likely possible to use linear symmetry at finer scales to facilitate quantisation of spots.

References

1. O'Farrell, P.H., High resolution two-dimensional electrophoresis of proteins, *J. Biol. Chem.* 250:4007-4021, 1975
2. Bigün, J., Bigün T., Symmetry derivatives of Gaussians illustrated by cross tracking, Research Report IDE-0131, 2001
3. Bigün, J., Optimal orientation detection of linear symmetry, IEEE Computer Society Press, Washington, DC, pages 433-438, 1987, In First International Conference on Computer Vision, ICCV (London).
4. Bigün, J., A structure feature for some image processing applications based on spiral functions, *Computer Vision, Graphics, and Image Processing*, 51(2):166-194, 1990
5. Raman, B., Cheung, A., Marten, M. R., Quantitative comparison and evaluation of two commercially available, two-dimensional electrophoresis image analysis software packages, Z3 and Melanie, *Electrophoresis* 23:2194-2202
6. http://userpages.umbc.edu/~braman1/proteome_index.html